UNIVERSITI TUNKU ABDUL RAHMAN

Department of Mathematics and Actuarial Science

CONTENTS

2 Linear Models

2.1 General Linear Models

Any linear model can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

$\uparrow$	$\uparrow$	$\uparrow$
observed responses	the elements of $\mathbf{X}$ are known (non-random) values	random errors are not observed

For the i-th case, the observed values are

$$(y_i \quad x_{i1} \quad x_{i2} \quad \cdots \quad x_{ik})$$

$\uparrow$	$\uparrow$
response variable	explanatory variables that describe conditions under which the response was generated.

where $\boldsymbol{\epsilon}$ specifying the distribution of the random error vector completes the specification of the distribution of $\mathbf{y}$

Note:

$$\boldsymbol{\epsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta} = \mathbf{y} - E(\mathbf{y})$$

Then,

$$E(\boldsymbol{\epsilon}) = \mathbf{0}$$
$$V(\boldsymbol{\epsilon}) = V(\mathbf{y}) = \boldsymbol{\Sigma}$$

Example 1. Regression Analysis: Yield of a chemical process

Yield (%)	Temperature (^oF)	Time (hr)
y	x_1	x_2
77	160	1
82	165	3
84	165	2
89	170	1
94	175	2

Simple linear regression model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$$
$$i = 1, 2, 3, 4, 5$$

Matrix formulation:

Example 2.
Blood coagulation times (in seconds) for blood samples from six different rats. Each rat was fed one of three diets.

Diet 1	Diet 2	Diet 3
$y_{11} = 62$	$y_{21} = 71$	$y_{31} = 72$
$y_{12} = 60$		$y_{32} = 68$
		$y_{33} = 67$

A **"means"** model

$$y_{ij} = \mu_i + \epsilon_{ij}$$

observed time for the j-th rat fed the i-th diet	mean time for rats given the i-th diet	random error with $E(\epsilon_{ij}) = 0$

You can express this model as

An **"effects"** model

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

This can be expressed as

This is a linear model with

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} \text{ and } V(\mathbf{y}) = \Sigma$$

You could add the assumptions

- independent errors
- homogeneous variance, i.e. $V(\epsilon_{ij}) = \sigma^2$

to obtain a linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ with

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} \text{ and } V(\mathbf{y}) = \sigma^2\mathbf{I}$$

Example 3. A 2×2 factorial experiment

- Experimental units: 8 plots with 5 trees per plot.
- Factor 1: Variety (A or B)
- Factor 2: Fungicide use (new or old)
- Response: Percentage of apples with spots

Percentage of apples with spots	Variety	Fungicide use
$y_{111} = 4.6$	A	new
$y_{112} = 7.4$	A	new
$y_{121} = 18.3$	A	old
$y_{122} = 15.7$	A	old
$y_{211} = 9.8$	B	new
$y_{212} = 14.2$	B	new
$y_{211} = 21.1$	B	old
$y_{222} = 18.9$	B	old

y_{ijk}	$= \mu + V_i$	$+ F_j$	$+ VF_{ij}$	$+ \epsilon_{ijk}$
↑	↑	↑	↑	↑
percent with spots	variety effects (i=1,2)	fung. use (j=1,2)	inter-action (k=1,2)	random error

Here we are using 9 parameters

$$\boldsymbol{\beta}^T = (\mu\ V_1\ V_2\ F_1\ F_2\ VF_{11}\ VF_{12}\ VF_{21}\ VF_{22})$$

to represent the 4 response means,

$$E(y_{ijk}) = \mu_{ij}, \quad i = 1, 2, \text{ and } j = 1, 2,$$

corresponding to the 4 combinations of levels of the two factors.

Write this model in the form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

A **"means"** model

$$y_{ijk} = \mu_{ij} + \epsilon_{ijk}$$

where
$\mu_{ij} = E(y_{ijk})$ = mean percentage of apples with spots. This linear model can be written in the form $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, that is,

The "effects" linear model and the "means" linear model are equivalent in the sense that the space of possible mean vectors

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$$

is the same for the two models.

- the model matrices differ
- the parameter vectors differ
- the columns of the model matrices span the same vector space

$$\begin{aligned} E(\mathbf{y}) &= \mathbf{X}\boldsymbol{\beta} \\ &= \left[\mathbf{x}_1 \middle| \mathbf{x}_2 \middle| \cdots \middle| \mathbf{x}_k\right] \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \\ &= \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \cdots + \beta_k \mathbf{x}_k \end{aligned}$$

2.2 Gauss-Markov Model

Definition 1.
The linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

is a **Gauss-Markov model** if

$$V(\mathbf{y}) = V(\boldsymbol{\epsilon}) = \sigma^2 I$$

for an unknown constant $\sigma^2 > 0$.

Notation: $$\mathbf{y} \underset{\substack{\text{distributed}\\ \text{as}}}{\sim} (\underset{E(\mathbf{y})}{\mathbf{X}\boldsymbol{\beta}},\ \underset{V(\mathbf{y})}{\sigma^2 I})$$

The distribution of $\mathbf{y}$ is not completely specified.

2.3 Normal Theory Gauss-Markov Model

Definition 2.
A normal-theory Gauss-Markov model is a Gauss-Markov model in which $\mathbf{y}$ (or $\boldsymbol{\epsilon}$) has a multivariate normal distribution.

$$\mathbf{y} \underset{\substack{\text{distr.}\\ \text{as}}}{\sim} \underset{\substack{\text{multivar.}\\ \text{normal}\\ \text{distr.}}}{N} (\underset{E(\mathbf{y})}{\mathbf{X}\boldsymbol{\beta}},\ \underset{V(\mathbf{y})}{\sigma^2 I})$$

The additional assumption of a normal distribution is

- not needed for some estimation results
- useful in creating
  - confidence intervals
  - tests of hypotheses

2.4 Ordinary Least Squares Estimation

For the linear model with

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} \text{ and } V(\mathbf{y}) = \boldsymbol{\Sigma}$$

we have

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1k} \\ X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{nk} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

and

$$\begin{aligned} y_i &= \boldsymbol{\beta}_1 \mathbf{x}_{i1} + \boldsymbol{\beta}_2 \mathbf{x}_{i2} + \cdots + \boldsymbol{\beta}_k \mathbf{x}_{ik} + \boldsymbol{\epsilon}_i \\ &= \mathbf{X}_i^T \boldsymbol{\beta} + \boldsymbol{\epsilon}_i \end{aligned}$$

where $\mathbf{X}_i^T = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \cdots, \mathbf{x}_{ik})$ is the i-th row of the model matrix $\mathbf{X}$.

Definition 3.
For a linear model with $E(\mathbf{y}) = X\boldsymbol{\beta}$, any vector $\mathbf{b}$ that minimizes the sum of squared residuals

$$\begin{aligned} Q(\mathbf{b}) &= \textstyle\sum_{i=1}^n (y_i - \mathbf{X}_i^T \mathbf{b})^2 \\ &= (\mathbf{y} - X\mathbf{b})^T (\mathbf{y} - \mathbf{X}\mathbf{b}) \end{aligned}$$

is an ordinary least squares (OLS) estimator for $\boldsymbol{\beta}$.

For $j = 1, 2, \ldots, k$, solve

$$0 = \frac{\partial Q(\mathbf{b})}{\partial b_j} = 2 \sum_{i=1}^n (y_i - \mathbf{X}_i^T \mathbf{b})\, X_{ij}$$

Dividing by 2, we have

$$0 = \sum_{i=1}^n (y_i - \mathbf{X}_i^T \mathbf{b})\, X_{ij} \quad j = 1, 2, \ldots, k$$

These equations are expressed in matrix form as

$$\begin{aligned} \mathbf{0} &= \mathbf{X}^T (\mathbf{y} - \mathbf{X}\mathbf{b}) \\ &= \mathbf{X}^T \mathbf{y} - \mathbf{X}^T \mathbf{X}\mathbf{b} \end{aligned}$$

or

$$\mathbf{X}^T \mathbf{X}\mathbf{b} = \mathbf{X}^T \mathbf{y}$$

These are often called the "normal" equations.

If $\mathbf{X}_{n\times k}$ has full column rank, i.e., rank$(\mathbf{X}) = k$, then

- $\mathbf{X}^T\mathbf{X}$ is non-singular
- $(\mathbf{X}^T\mathbf{X})^{-1}$ exists and is unique

Consequently,

$$(\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{X})\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

and

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

is the unique solution to the normal equations.

If rank$(\mathbf{X}) < k$, then

- there are infinitely many solutions to the normal equations
- if $\mathbf{G} = (\mathbf{X}^T\mathbf{X})^-$ is a generalized inverse of $\mathbf{X}^T\mathbf{X}$, then

  $$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^- \mathbf{X}^T\mathbf{y} = \mathbf{G}\mathbf{X}^T\mathbf{y}$$

  is a solution of the normal equations.

2.5 Generalized Inverse

Definition 4.
For a given $m \times n$ matrix $\mathbf{A}$, any $n \times m$ matrix $\mathbf{G}$ that satisfies

$$\mathbf{AGA} = \mathbf{A}$$

is a **generalized inverse** of $\mathbf{A}$.

Comments:

(i) We will often use $\mathbf{A}^-$ to denote a generalized inverse of $\mathbf{A}$.

(ii) There may be infinitely many generalized inverses.

(iii) If $\mathbf{A}$ is an $m \times m$ nonsingular matrix, then $\mathbf{G} = \mathbf{A}^{-1}$ is the unique generalized inverse for $\mathbf{A}$.

Example 4.

$$\mathbf{A} = \begin{bmatrix} 16 & -6 & -10 \\ -6 & 21 & -15 \\ -10 & -15 & 25 \end{bmatrix} \text{ with } \operatorname{rank}(\mathbf{A}) = 2.$$

Check that

$$\mathbf{G}_1 = \begin{bmatrix} \frac{1}{20} & 0 & 0 \\ 0 & \frac{1}{30} & 0 \\ 0 & 0 & \frac{1}{50} \end{bmatrix} \text{ and } \mathbf{G}_2 = \begin{bmatrix} \frac{21}{300} & \frac{6}{300} & 0 \\ \frac{6}{300} & \frac{16}{300} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

are generalized inverse of $\mathbf{A}$.

Example 5.

Show tthat if $\mathbf{X}_{n\times k}$ has $\operatorname{rank}(\mathbf{X}) < k$,and if $\mathbf{G} = (\mathbf{X}^T\mathbf{X})^-$ is a generalized inverse of $\mathbf{X}^T\mathbf{X}$, then

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^- \mathbf{X}^T\mathbf{y} = \mathbf{G}\mathbf{X}^T\mathbf{y}$$

is a solution of the normal equations.

Example 6.

A "means" model is as follow:

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{31} \\ y_{32} \\ y_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \end{bmatrix}$$

(i) Compute $\mathbf{X}^T\mathbf{X}$ and $\mathbf{X}^T\mathbf{y}$.

(ii) Obtain the OLS estimator.

Example 7.

"Effects" model

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$
$$i = 1, 2, 3; j = 1, 2, \ldots, n_i$$

(i) Write out the $\mathbf{X}^T\mathbf{X}$ matrix for this models.

(ii) Check that $(\mathbf{X}^T\mathbf{X})^- = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{n_1} & 0 & 0 \\ 0 & 0 & \frac{1}{n_2} & 0 \\ 0 & 0 & 0 & \frac{1}{n_3} \end{bmatrix}$ is a generalized inversed of $\mathbf{X}^T\mathbf{X}$ and compute the corresponding solution to the normal equations.

2.5.1 Evaluating Generalized Inverses

Step(1) Find any $r \times r$ nonsingular submatrix of $\mathbf{A}$ where r=rank($\mathbf{A}$). Call this matrix $\mathbf{W}$.

Step(2) Invert and transpose $\mathbf{W}$, ie., compute $(\mathbf{W}^{-1})^T$.

Step(3) Replace each element of $\mathbf{W}$ in $\mathbf{A}$ with the corresponding element of $(\mathbf{W}^{-1})^T$

Step(4) Replace all other elements in $\mathbf{A}$ with zeros.

Step(5) Transpose the resulting matrix to obtain $\mathbf{G}$, a generalized inverse for $\mathbf{A}$.

Example 8.

$$\mathbf{A} = \begin{bmatrix} 4 & 1 & 2 & 0 \\ 1 & 1 & 5 & 15 \\ 3 & 1 & 3 & 5 \end{bmatrix}$$

$$\mathbf{W} = \begin{bmatrix} 1 & 5 \\ 1 & 3 \end{bmatrix}$$

You are given that $rank(\mathbf{A}) = 2$, find $\mathbf{G} = \mathbf{A}^-$.

Example 9.

$$\mathbf{A} = \begin{bmatrix} 4 & 1 & 2 & 0 \\ 1 & 1 & 5 & 15 \\ 3 & 1 & 3 & 5 \end{bmatrix}$$

$$\mathbf{W} = \begin{bmatrix} 4 & 0 \\ 3 & 5 \end{bmatrix}$$

You are given that $rank(\mathbf{A}) = 2$, find $\mathbf{G} = \mathbf{A}^{-}$.

Example 10.

In an experiment to investigate the effect of nitrogen fertilizer on lettuce production. Two rates of ammonium were applied to 5 replicates plots in a completely randomized design. The data are the number of heads of lettuce harvested from the plot.

		j				
i	Treatment(lb N/acre)	1	2	3	4	5
1	0	104	114	90	140	135
2	50	134	130	144	174	189

Consider the linear model

$y_{ij} = \mu + \tau_i + \epsilon_{ij}$, for $i = 1, 2$ and $j = 1, 2, \ldots, 5$

where

- y_{ij} is the observed number of heads of lettuce for the i^{th} fertilizer assigned to the j^{th} plot.
- τ_i corresponds to the effect of i^{th} fertilizer.
- $\epsilon_{ij} \sim N(0, \sigma^2)$.

(a) Write model above in the form $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$. Do not impose any restriction on the parameters.

(b) Obtain two generalized inverses of $\mathbf{X^T X}$, $\mathbf{G_1}$ and $\mathbf{G_2}$.

(c) Use the generalized inverse you obtained in part(b) to compute solutions to the normal equations, $\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\mu} \\ \hat{\tau}_1 \\ \hat{\tau}_2 \end{bmatrix}$.

(d) Using your solution $\hat{\boldsymbol{\beta}}$ to the normal equation from part (c), estimates $\gamma_1 = \tau_1 - \tau_2$.

(e) Using your solutions $\hat{\boldsymbol{\beta}}$ to the normal equation from part (c), estimates $\gamma_2 = \tau_1 + \tau_2$.

2.5.2 Moore-Penrose Inverse

Definition 5. For any matrix $\mathbf{A}$ there is a **unique** matrix M, called the Moore-Penrose inverse, that satisfies

(i) $\mathbf{A}M\mathbf{A} = \mathbf{A}$

(ii) $M\mathbf{A}M = M$

(iii) $\mathbf{A}M$ is symmetric

(iv) $M\mathbf{A}$ is symmetric

2.5.3 Properties of generalized inverses of $\mathbf{X}^T\mathbf{X}$

Result 1. If $\mathbf{G}$ is a generalized inverse of $\mathbf{X}^T\mathbf{X}$, then

(i) $\mathbf{G}^T$ is a generalized inverse of $\mathbf{X}^T\mathbf{X}$.

(ii) $\mathbf{X}\mathbf{G}\mathbf{X}^T\mathbf{X} = \mathbf{X}$, i.e., $\mathbf{G}\mathbf{X}^T$ is a generalized inverse of $\mathbf{X}$.

(iii) $\mathbf{X}\mathbf{G}\mathbf{X}^T$ is invariant with respect to the choice of $\mathbf{G}$.

(iv) $\mathbf{X}\mathbf{G}\mathbf{X}^T$ is symmetric.

.

.

2.6 Estimation of the Mean Vector

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$$

For any solution to the normal equations, say

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T\mathbf{y} \ ,$$

2.6.1 OLS Estimator $E(\mathbf{y})$

The OLS estimator for $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ is

$$\begin{aligned}\hat{\mathbf{y}} &= \mathbf{X}\mathbf{b}\\ &= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T\mathbf{y}\\ &= P_{\mathbf{X}}\mathbf{y}\end{aligned}$$

- The matrix $P_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T$ is called an "orthogonal projection matrix".
- $\hat{\mathbf{y}} = P_{\mathbf{X}}\mathbf{y}$ is the projection of $\mathbf{y}$ onto the space spanned by the columns of $\mathbf{X}$.

Result 2. Properties of a projection matrix

$$P_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T$$

(i) $P_{\mathbf{X}}$ is invariant to the choice of $(\mathbf{X}^T\mathbf{X})^-$ For any solution $\mathbf{b} = (\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{y}$ to the normal equations

$$\hat{\mathbf{y}} = \mathbf{X}\mathbf{b} = P_{\mathbf{X}}\mathbf{y}$$

is the same. (From Result 1(iii))

(ii) $P_{\mathbf{X}}$ is symmetric (From Result 1 (iv))

(iii) $P_{\mathbf{X}}$ is idempotent ($P_X P_X = P_X$)

(iv) $P_X\mathbf{X} = \mathbf{X}$ (From Result 1 (ii))

(v) Partition $\mathbf{X}$ as $\mathbf{X} = [\mathbf{X}_1|\mathbf{X}_2|\cdots|\mathbf{X}_k]$, then $P_{\mathbf{X}}\mathbf{X}_j = \mathbf{X}_j$

2.6.2 Residuals

$$\mathbf{e}_i = \mathbf{y}_i - \hat{\mathbf{y}}_i \quad i = 1, \ldots, n$$

$$\begin{aligned}\mathbf{e} &= \mathbf{y} - \hat{\mathbf{y}}\\ &= \mathbf{y} - \mathbf{X}\mathbf{b}\\ &= \mathbf{y} - \mathbf{P}_{\mathbf{X}}\mathbf{y}\\ &= (\mathbf{I} - \mathbf{P}_{\mathbf{X}})\mathbf{y}\end{aligned}$$

Comment: $\mathbf{I} - \mathbf{P}_{\mathbf{X}}$ is a projection matrix that projects $\mathbf{y}$ onto the space orthogonal to the space spanned by the columns of $\mathbf{X}$.

Result 3.

(i) $\mathbf{I} - \mathbf{P}_{\mathbf{X}}$ is symmetric

(ii) $\mathbf{I} - \mathbf{P}_{\mathbf{X}}$ is idempotent

$$(\mathbf{I} - \mathbf{P}_{\mathbf{X}})(\mathbf{I} - \mathbf{P}_{\mathbf{X}}) = \mathbf{I} - \mathbf{P}_{\mathbf{X}}$$

(iii) $(\mathbf{I} - \mathbf{P}_{\mathbf{X}})\mathbf{P}_{\mathbf{X}} = 0$

(iv) $(\mathbf{I} - \mathbf{P}_{\mathbf{X}})\mathbf{X} = \mathbf{0}$

(v) Partition $\mathbf{X}$ as $[\mathbf{x}_1|\mathbf{x}_2|\cdots|\mathbf{x}_k]$ then

$$(\mathbf{I} - \mathbf{P}_{\mathbf{X}})\mathbf{X}_j = \mathbf{0}$$

(vi) Residuals are invariant with respect to the choice of $(\mathbf{X}^T\mathbf{X})^-$, so

$$\mathbf{e} - \mathbf{y} - \mathbf{X}\mathbf{b} = (\mathbf{I} - \mathbf{P}_{\mathbf{X}})\mathbf{y}$$

is the same for any solution

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{y}$$

to the normal equations

The residual vector

$$\mathbf{e} = \mathbf{y} - \tilde{\mathbf{y}} = (\mathbf{I} - \mathbf{P}_{\mathbf{X}})\mathbf{y}$$

is in the space orthogonal to the space spanned by the columns of $\mathbf{X}$. It has dimension

$$n - rank(\mathbf{X}).$$

2.6.3 Partition of a total sum of squares

Squared length of $\mathbf{y}$ is

$$\sum_{i=1}^{n} y_i^2 = \mathbf{y}^T\mathbf{y}$$

Squared length of the residual vector is

$$\begin{aligned}\sum_{i=1}^{n} e_i^2 &= \mathbf{e}^T\mathbf{e}\\ &= [(I - P_X)\mathbf{y}]^T(I - P_X)\mathbf{y}\\ &= \mathbf{y}^T(I - P_X)\mathbf{y}\end{aligned}$$

Squared length of $\hat{\mathbf{y}} = P_X\mathbf{y}$ is

$$\begin{aligned}\sum_{i=1}^{n} \hat{y}_i^2 &= \hat{\mathbf{y}}^T\hat{\mathbf{y}}\\ &= (P_X\mathbf{y})^T(P_X\mathbf{y})\\ &= \mathbf{y}^T(P_X)^T P_X\mathbf{y} \quad \text{since } P_X \text{ is symmetric}\\ &= \mathbf{y}^T P_X P_X\mathbf{y} \quad \text{since } P_X \text{ is idempotent}\\ &= \mathbf{y}^T P_X\mathbf{y}\end{aligned}$$

We have

$$\mathbf{y}^T\mathbf{y} = \mathbf{y}^T(P_X + I - P_X)\mathbf{y}$$
$$= \mathbf{y}^T P_X \mathbf{y} + \mathbf{y}^T(I - P_X)\mathbf{y}.$$

Analysis of Variance Table

Source of Variation	Degrees of Freedom	Sums of Squares
model (uncorrected)	rank($\mathbf{X}$)	$\hat{\mathbf{y}}^T\hat{\mathbf{y}} = \mathbf{y}^T P_X \mathbf{y}$
residuals	n-rank($\mathbf{X}$)	$\mathbf{e}^T\mathbf{e} = \mathbf{y}^T(I - P_X)\mathbf{y}$
total (uncorrected)	n	$\mathbf{y}^T\mathbf{y} = \sum_{i=1}^{n} y_i^2$

Result 4.
For the linear model with

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} \text{ and } V(y) = \boldsymbol{\Sigma},$$

the OLS estimator

$$\hat{\mathbf{y}} = \mathbf{X}\mathbf{b} = P_X\mathbf{y}$$

for

$$\mathbf{X}\boldsymbol{\beta}$$

is

(i) unbiased, i.e., $E(\hat{\mathbf{y}}) = \mathbf{X}\boldsymbol{\beta}$

(ii) a linear function of $\mathbf{y}$

(iii) has variance-covariance matrix

$$V(\hat{\mathbf{y}}) = P_X \Sigma P_X$$

This is true for any solution

$$b = (\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T\mathbf{y}$$

to the normal equations.

Comments:

(i) $\hat{\mathbf{y}} = \mathbf{X}\mathbf{b} = P_X\mathbf{y}$ is said to be a **linear unbiased** estimator for $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$

(ii) For the Gauss-Markov model, $V(\mathbf{y}) = \sigma^2 I$ and

$$V(\hat{\mathbf{y}}) = P_X(\sigma^2 I)P_X$$
$$= \sigma^2 P_X P_X$$
$$= \sigma^2 P_X$$
$$= \sigma^2 \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T$$

$\uparrow$ this is sometimes called the “hat” matrix.

2.7 Estimable Functions

Some estimates of linear functions of the parameters have the same value, regardless of which solution to the normal equations is used. These are called estimable functions.

Definition 6.
For a linear model with

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} \quad \text{and} \quad V(\mathbf{y}) = \boldsymbol{\Sigma}$$

we will say that

$$\mathbf{c}^T\boldsymbol{\beta} = c_1\beta_1 + c_2\beta_2 + \cdots + c_k\beta_k$$

is **estimable** if there exists a linear unbiased estimator $\mathbf{a}^T\mathbf{y}$ for $\mathbf{c}^T\boldsymbol{\beta}$, i.e., for some vector of constants $\mathbf{a}$, we have $E(\mathbf{a}^T\mathbf{y}) = \mathbf{c}^T\boldsymbol{\beta}$.

We will use **Blood coagulation times** example to illustrate estimable and non-estimable functions.

Diet 1	Diet 2	Diet 3
$y_{11} = 62$	$y_{21} = 71$	$y_{31} = 72$
$y_{12} = 60$		$y_{32} = 68$
		$y_{33} = 67$

The "Effects" model

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

can be written as

$$\underset{\mathbf{y}}{\begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{31} \\ y_{32} \\ y_{33} \end{bmatrix}} \underset{X}{\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}} \underset{\boldsymbol{\beta}}{\begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}} + \underset{\boldsymbol{\epsilon}}{\begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \end{bmatrix}}$$

Note that

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} \quad \text{or} \quad E(\boldsymbol{\epsilon}) = 0$$

2.7.1 Example of Estimable Functions

Example 11.
Show that $\mu + \alpha_1$ is estimable.

Example 12.
Show that $\mu + \alpha_2$ is estimable.

Example 13.
Show that $\mu + \alpha_3$ is estimable.

Example 14.
Show that $\alpha_1 - \alpha_2$ is estimable.

Example 15.
Show that $2\mu + 3\alpha_1 - \alpha_2$ is estimable.

2.7.2 Quantities that are not estimable

Quantities that are **not** estimable include

$$\mu, \alpha_1, \alpha_2, \alpha_3, 3\alpha_1, 2\alpha_2$$

To show that a linear function of parameters

$$c_0\mu + c_1\alpha_1 + c_2\alpha_+ c_3\alpha_3$$

is not estimable, one must show that there is no non-random vector

$$\mathbf{a}^{\mathbf{T}} = (a_0, a_1, a_2, a_3)$$

For which

$$E(\mathbf{a}^{\mathbf{T}}\mathbf{y}) = c_0\mu + c_1\alpha_1 + c_2\alpha_2 + c_3\alpha_3$$

Example 16.
Show that α_1 is not estimable.

Result 5.

For a model with $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ and $V(y) = \boldsymbol{\Sigma}$:

(i) The expectation of any observation is estimable.

(ii) A linear combination of estimable functions is estimable.

(iii) Each element of $\boldsymbol{\beta}$ is estimable if and only if $\text{rank}(\mathbf{X}) = k$ = number of columns.

(iv) Every $\mathbf{c}^T\boldsymbol{\beta}$ is estimable if and only if $\text{rank}(\mathbf{X}) = k$ = number of columns in $\mathbf{X}$.

.

.

Result 6. For a linear model with $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ and $V(\mathbf{y}) = \Sigma$, each of the following is true if and only if $\mathbf{c}^T\boldsymbol{\beta}$ is **estimable**.

(i) $\mathbf{c}^T = \mathbf{a}^T\mathbf{X}$ for some $\mathbf{a}$ i.e., $\mathbf{c}$ is in the space spanned by the rows of $\mathbf{X}$.

(ii) $\mathbf{c}^T\mathbf{a} = 0$ for every $\mathbf{a}$ for which $\mathbf{X}\mathbf{a} = \mathbf{0}$.

(iii) $\mathbf{c}^T\mathbf{b}$ is the same for any solution to the normal equations $(\mathbf{X}^T\mathbf{X})\mathbf{b} = \mathbf{X}^T\mathbf{y}$, i.e., there is a **unique** least squares estimator for $\mathbf{c}^T\boldsymbol{\beta}$.

.

Example 17.
Use Result 6 (ii) to show that μ is not estimable.

Part (ii) of Result 6 sometimes provides a convenient way to identify all possible estimable functions of $\boldsymbol{\beta}$.
In Blood Coagulation Times example,

$$\mathbf{Xd} = \mathbf{0}$$

if and only if

$$\mathbf{d} = w \begin{bmatrix} 1 \\ -1 \\ -1 \\ -1 \end{bmatrix}$$

for some scalar w.

Then,

$$\mathbf{c}^T\boldsymbol{\beta}$$

is estimable if and only if

$$0 = \mathbf{c}^T\mathbf{d} = w(c_1 - c_2 - c_3 - c_4) = 0$$
$$\iff \quad c_1 = c_2 + c_3 + c_4.$$

Then,

$(c_2 + c_3 + c_4)\mu + c_2\alpha_1 + c_3\alpha_2 + c_4\alpha_3$

is estimable for any $(c_2\ c_3\ c_4)$ and these are the only estimable functions of $\mu, \alpha_1, \alpha_2, \alpha_3$.

For example, some estimable functions are

$$\mu + \frac{1}{3}(\alpha_1 + \alpha_2 + \alpha_3) \;\; (c_2 = c_3 = c_4 = \frac{1}{3})$$

and

$$\mu + \alpha_2 \quad (c_2 = 1\ c_3 = c_4 = 0)$$

but

$$\mu + 2\alpha_2$$

is not estimable

Example 18.
Let

$$\mathbf{y} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \end{bmatrix}$$

Show that every linear parametric function $c_1\beta_1 + c_2\beta_2$ is estimable.

Definition 7.
For a linear model with $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ and $V(\mathbf{y}) = \Sigma$, where $\mathbf{X}$ is an $n \times k$ matrix, $C_{r\times k}\boldsymbol{\beta}_{k\times 1}$ is said to be **estimable** if all of its elements

$$C\boldsymbol{\beta} = \begin{bmatrix} \mathbf{c}_1^T \\ \mathbf{c}_2^T \\ \vdots \\ \mathbf{c}_r^T \end{bmatrix} \boldsymbol{\beta} = \begin{bmatrix} \mathbf{c}_1^T\boldsymbol{\beta} \\ \mathbf{c}_2^T\boldsymbol{\beta} \\ \vdots \\ \mathbf{c}_r^T\boldsymbol{\beta} \end{bmatrix}$$

are estimable.

Result 7.
For the linear model with $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ and $V(\mathbf{y}) = \Sigma$, where $\mathbf{X}$ is an $n\times k$ matrix. Each of the following conditions hold if and only if $C\boldsymbol{\beta}$ is estimable.

(i) $\mathbf{A}^T\mathbf{X} = C$ for some matrix $\mathbf{A}$, i.e., each row of C is in the space spanned by the rows of $\mathbf{X}$.

(ii) $C\mathbf{d} = \mathbf{0}$ for any $\mathbf{d}$ for which $\mathbf{X}\mathbf{d} = \mathbf{0}$.

(iii) $C\mathbf{b}$ is the same for any solution to the normal equations $(\mathbf{X}^T\mathbf{X})\mathbf{b} = \mathbf{X}^T\mathbf{y}$.

2.8 Best Linear Unbiased Estimator

For a linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

with

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} \text{ and } V(\mathbf{y}) = \boldsymbol{\Sigma},$$

we have

- Any estimable function has a specific interpretation
- The OLS estimator for an estimable function $C\boldsymbol{\beta}$ is unique
$$C\mathbf{b} = C\,(\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T\mathbf{y}$$
- The OLS estimator for an estimable function $C\boldsymbol{\beta}$ is
  - a linear estimator
  - an unbiased estimator

In the class of linear unbiased estimators for $\mathbf{c}^T\boldsymbol{\beta}$, is the OLS estimator the “best?”

Here "best" means smallest expected squared error. Let $t(\mathbf{y})$ denote a linear unbiased estimator for $\mathbf{c}^T\boldsymbol{\beta}$. Then, the expected squared error is

$$\begin{aligned}
\text{MSE} &= E[t(\mathbf{y}) - \mathbf{c}^T\boldsymbol{\beta}]^2 \\
&= E[t(\mathbf{y}) - E(t(\mathbf{y})) + E(t(\mathbf{y})) - \mathbf{c}^T\boldsymbol{\beta}]^2 \\
&= E[t(\mathbf{y}) - E(t(\mathbf{y}))]^2 \\
&\quad + [E(t(\mathbf{y})) - \mathbf{c}^T\boldsymbol{\beta}]^2 \\
&\quad + 2[E(t(\mathbf{y})) - \mathbf{c}^T\boldsymbol{\beta}]E[t(\mathbf{y}) - E(t(\mathbf{y}))] \\
&= E[t(\mathbf{y}) - E(t(\mathbf{y}))]^2 + [E(t(\mathbf{y})) - \mathbf{c}^T\boldsymbol{\beta}]^2 \\
&= V(t(\mathbf{y})) + [\text{bias}]^2
\end{aligned}$$

$\uparrow$

the bias is zero

We are restricting our attention to linear unbiased estimators for $\mathbf{c}^T\boldsymbol{\beta}$:

- $E(t(\mathbf{y})) = \mathbf{c}^T\boldsymbol{\beta}$
- $t(\mathbf{y}) = \mathbf{a}^T\mathbf{y}$ for some $\mathbf{a}$

Then, $t(\mathbf{y}) = \mathbf{a}^T\mathbf{y}$ is the **Best Linear Unbiased Estimator (BLUE)** for $\mathbf{c}^T\boldsymbol{\beta}$ if

$$V(\mathbf{a}^T\mathbf{y}) \leq V(\mathbf{d}^T\mathbf{y})$$

for any $\mathbf{d}$ and any value of $\boldsymbol{\beta}$.

Result 8. Gauss-Markov Theorem
For the Gauss-Markov model with

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} \text{ and } V(\mathbf{y}) = \sigma^2 I$$

the OLS estimator of an estimable function $\mathbf{c}^T\boldsymbol{\beta}$ is the **unique** best linear unbiased estimator (blue) of $\mathbf{c}^T\boldsymbol{\beta}$.

.

.

.

.

2.9 Generalized Least Squares (GLS) Estimation

What if you have a linear model that is **not** a Gauss-Markov model?

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$$
$$V(\mathbf{y}) = \Sigma \neq \sigma^2 I$$

- Parts (i) and (ii) of the proof of result 8 do not require
$$V(\mathbf{y}) = \sigma^2 I \ .$$

  Consequently, the OLS estimator for $\mathbf{c}^T\boldsymbol{\beta}$,
$$\mathbf{c}^T\mathbf{b} = \mathbf{c}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{y}$$
  is a linear unbiased estimator.

- Result 6 does not require
$$V(\mathbf{y}) = \sigma^2 I$$
  and the OLS estimator for any estimable quantity,
$$\mathbf{c}^T\mathbf{b} = \mathbf{c}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{y}\,,$$
  is invariant to the choice of $(\mathbf{X}^T\mathbf{X})^-$.

- The OLS estimator $\mathbf{c}^T\mathbf{b}$ may not be blue. There may be other linear unbiased estimators with smaller variance.

Note

$$\begin{aligned} V(\mathbf{c}^T\mathbf{b}) &= V(\mathbf{c}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{y}) \\ &= \mathbf{c}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\Sigma\mathbf{X}[(\mathbf{X}^T\mathbf{X})^-]^T\mathbf{c} \end{aligned}$$

For the Gauss-Markov model

$$V(\mathbf{y}) = \Sigma = \sigma^2 I$$

and

$$\begin{aligned} V(\mathbf{c}^T\mathbf{b}) &= \sigma^2\mathbf{c}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{X}[(\mathbf{X}^T\mathbf{X})^-]^T\mathbf{c} \\ &= \sigma^2\mathbf{c}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{c} \end{aligned}$$

$$\begin{aligned} E(\mathbf{y}) &= \mathbf{X}\boldsymbol{\beta} \\ V(\mathbf{y}) &= \Sigma \neq \sigma^2 I \end{aligned}$$

Definition 8.
For a linear model with

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} \ \text{ and } \ V(\mathbf{y}) = \Sigma,$$

where Σ is positive definite, a generalized least squares estimator for $\boldsymbol{\beta}$ minimizes

$$(\mathbf{y} - \mathbf{X}\mathbf{b}_{\text{GLS}})^T\Sigma^{-1}(\mathbf{y} - \mathbf{X}\mathbf{b}_{\text{GLS}})$$

Strategy: Transform $\mathbf{y}$ to a random vector $\mathbf{Z}$ for which the Gauss-Markov model applies.
The spectral decomposition of Σ yields

$$\Sigma = \sum_{j=1}^{n} \lambda_j \mathbf{u}_j\mathbf{u}_j^T.$$

Define

$$\Sigma^{-1/2} = \sum_{j=1}^{n} \frac{1}{\sqrt{\lambda_j}}\mathbf{u}_j\mathbf{u}_j^T$$

and create the random vector $\mathbf{Z} = \Sigma^{-1/2}\mathbf{y}$.

Then

$$\begin{aligned} V(\mathbf{Z}) &= V(\Sigma^{-1/2}\mathbf{y}) \\ &= \Sigma^{-1/2}\Sigma\,\Sigma^{-1/2} \\ &= I \end{aligned}$$

and

$$\begin{aligned} E(\mathbf{Z}) &= E(\Sigma^{-1/2}\mathbf{y}) \\ &= \Sigma^{-1/2}E(\mathbf{y}) \\ &= \Sigma^{-1/2}\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{W}\boldsymbol{\beta} \end{aligned}$$

and we have a Gauss-Markov model for $\mathbf{Z}$, where $\mathbf{W} = \Sigma^{-1/2}\mathbf{X}$ is the model matrix.

Note that

$$\begin{aligned} &(\mathbf{Z} - \mathbf{W}\mathbf{b})^T(\mathbf{Z} - \mathbf{W}\mathbf{b}) \\ &= (\Sigma^{-1/2}\mathbf{y} - \Sigma^{1/2}\mathbf{X}\mathbf{b})^T(\Sigma^{-1/2}\mathbf{y}\Sigma^{-1/2}\mathbf{X}\mathbf{b}) \\ &= (y - \mathbf{X}\mathbf{b})^T\Sigma^{-1/2}\Sigma^{-1/2}(y - \mathbf{X}\mathbf{b}) \\ &= (y - \mathbf{X}\mathbf{b})^T\Sigma^{-1}(y - \mathbf{X}\mathbf{b}) \end{aligned}$$

Hence, any GLS estimator for the $\mathbf{y}$ model is an OLS estimator for the $\mathbf{Z}$ model.

It must be a solution to the normal equations for the $\mathbf{Z}$ model

$$\begin{aligned} &\mathbf{W}^T\mathbf{W}\mathbf{b} = \mathbf{W}^T\mathbf{Z} \\ &\Leftrightarrow (\mathbf{X}^T\Sigma^{-1/2}\Sigma^{-1/2}\mathbf{X})\mathbf{b} \\ &\qquad = \mathbf{X}^T\Sigma^{-1/2}\Sigma^{-1/2}\mathbf{y} \\ &\Leftrightarrow (\mathbf{X}^T\Sigma^{-1}\mathbf{X})\mathbf{b} = \mathbf{X}^T\Sigma^{-1}\mathbf{y} \end{aligned}$$

These are the generalized least squares estimating equations.
Any solution

$$\begin{aligned} \mathbf{b}_{\text{GLS}} &= (\mathbf{W}^T\mathbf{W})^-\mathbf{W}^T\mathbf{Z} \\ &= (\mathbf{X}^T\Sigma^{-1}\mathbf{X})^-\mathbf{X}^T\Sigma^{-1}\mathbf{y} \end{aligned}$$

is called a generalized least squares (GLS) estimator for $\boldsymbol{\beta}$.

Result 9.
For the linear model with $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ and $V(\mathbf{y}) = \Sigma$ the GLS estimator of an estimable function $\mathbf{c}^T\boldsymbol{\beta}$,

$$\mathbf{c}^T\mathbf{b}_{\text{GLS}} = \mathbf{c}^T(\mathbf{X}^T\Sigma^{-1}\mathbf{X})^-\mathbf{X}^T\Sigma^{-1}\mathbf{y} ,$$

is the unique blue of $\mathbf{c}^T\beta$.

Comments:

- For the linear model with
$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} \text{ and } V(\mathbf{y}) = \Sigma$$
both the OLS and GLS estimators for an estimable function $\mathbf{c}^T\boldsymbol{\beta}$ are linear unbiased estimators.

$$V(\mathbf{c}^T\mathbf{b}_{\text{OLS}}) = \mathbf{c}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\Sigma\mathbf{X}[(\mathbf{X}^T\mathbf{X})^-]^T\mathbf{c}$$

$$V(\mathbf{c}^T\mathbf{b}_{\text{GLS}}) = \mathbf{c}^T(\mathbf{X}^T\Sigma^{-1}\mathbf{X})^-\mathbf{X}^T\Sigma^{-1}\mathbf{X}(\mathbf{X}^T\Sigma^{-1}\mathbf{X})^-\mathbf{c}$$

$\mathbf{c}^T\mathbf{b}_{\text{OLS}}$ is not a "bad" estimator, but
$$V(\mathbf{c}^T\mathbf{b}_{\text{OLS}}) \geq V(\mathbf{c}^T\mathbf{b}_{\text{GLS}})$$
because $\mathbf{c}^T\mathbf{b}_{\text{GLS}}$ is the unique blue for $\mathbf{c}^T\boldsymbol{\beta}$.

- For the Gauss-Markov model,
$$\mathbf{c}^T b_{\text{GLS}} = \mathbf{c}^T b_{\text{OLS}} .$$

- The results for $\mathbf{b}_{\text{GLS}}$ and $\mathbf{c}^T\mathbf{b}_{\text{GLS}}$ assume that $V(\mathbf{y}) = \boldsymbol{\Sigma}$ is known.

- The same results, hold for the model where $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ and $V(\mathbf{y}) = \sigma^2 V$ for some known matrix V.

- In practice $V(\mathbf{y}) = \boldsymbol{\Sigma}$ is usually unknown. Then an approximation to
$$\mathbf{b}_{\text{GLS}} = (\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^-\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{y}$$
is obtained by substituting a consistent

estimator $\hat{\boldsymbol{\Sigma}}$ for $\boldsymbol{\Sigma}$.

- use method of moments or maximum likelihood estimation to obtain $\hat{\boldsymbol{\Sigma}}$
- the resulting estimator
  * is not a linear estimator
  * is consistent but not necessarily unbiased
  * does not provide a blue for estimable functions
  * may have larger mean squared error than the OLS estimator

To create confidence intervals or test hypotheses about estimable functions for a linear model with
$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} \text{ and } V(\mathbf{y}) = \boldsymbol{\Sigma}$$
we must

(i) specify a probability distribution for $\mathbf{y}$ so we can derive a distribution for
$$\mathbf{c}^T\mathbf{b} = \mathbf{c}^T(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^-\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{y}$$

(ii) estimate σ^2 when
$$V(\mathbf{y}) = \sigma^2 I \text{ or } V(\mathbf{y}) = \sigma^2 V$$
for some known V.

(iii) Estimate $\boldsymbol{\Sigma}$ when
$$V(\mathbf{y}) = \boldsymbol{\Sigma}$$

Example 19.

Suppose that y_{11} and y_{12} are independent $N(\mu_1, 9\sigma^2)$ variables independent of y_{21} and y_{22} that are independent $N(\mu_2, 25\sigma^2)$ and $N(\mu_2, 36\sigma^2)$ variables respectively. What is the BLUE of $2\mu_1 + 3\mu_2$? Explain carefully.

.

.

Example 20.

Suppose $y_i = x_i\beta + \epsilon$, where for $\mathbf{e} = (\epsilon_1, \epsilon_2, \cdots, \epsilon_n)'$, $E(\mathbf{e}) = \mathbf{0}$. A particular experiment produces $n = 5$ data points as per

x	27	35	31	40	90
y	327	390	120	138	85

Suppose that $V(\epsilon) = \sigma^2 diag(9, 36, 64, 81, 100)$. Evaluate an appropriate BLUE of β under the model assumptions.

2.10 Reparameterization, Restrictions, and Avoiding Generalized Inverses

Models that may appear to be different at first sight, may be equivalent in many ways.

Example 21. Two-way classification
Consider the "cell mean" model.

$$y_{ijk} = \mu_{ij} + \epsilon_{ijk} \quad i = 1,2; j = 1,2; k = 1,2$$

where $\epsilon_{ijk} \sim NID(0, \sigma^2)$

$$\begin{bmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{21} \\ \mu_{22} \end{bmatrix} + \begin{bmatrix} \epsilon_{111} \\ \epsilon_{112} \\ \epsilon_{121} \\ \epsilon_{122} \\ \epsilon_{211} \\ \epsilon_{212} \\ \epsilon_{221} \\ \epsilon_{222} \end{bmatrix}$$

or

$$\mathbf{y} = \mathbf{W}\boldsymbol{\gamma} + \boldsymbol{\epsilon}$$

The "effects" model:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$$

where

$$\epsilon_{ijk} \sim NID(0, \sigma^2) \quad i = 1,2; j = 1,2; k = 1,2$$

$$\begin{bmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \gamma_{11} \\ \gamma_{12} \\ \gamma_{21} \\ \gamma_{22} \end{bmatrix}$$

or

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

The models are "equivalent": the space spanned by the columns of $\mathbf{W}$ is the same as the space spanned by columns of $\mathbf{X}$, i.e. $\mathcal{C}(\mathbf{W}) = \mathcal{C}(\mathbf{X})$.

You can find matrices F and $\mathbf{G}$ such that

$$\mathbf{W} = X \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = XF$$

and

$$X = \mathbf{W} \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} = \mathbf{W}\mathbf{G}$$

Then,

(i) rank(X) = rank $(\mathbf{W})$

(ii) Estimated mean responses are the same:

$$\begin{aligned} \hat{\mathbf{y}} &= X(X^TX)^-X^T\mathbf{y} \\ &= \mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{y} \end{aligned}$$

or

$$\hat{\mathbf{y}} = P_X\mathbf{y} = P_{\mathbf{W}}\mathbf{y}$$

(iii) Residual vectors are the same

$$\begin{aligned} \mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} &= (I - P_X)\mathbf{y} \\ &= (I - P_{\mathbf{W}})\mathbf{y} \end{aligned}$$

Example 22. Regression model for the yield of a chemical process

$$\underset{\text{yield}}{\underset{\uparrow}{y_i}} = \beta_0 + \beta_1 \underset{\text{temperature}}{\underset{\uparrow}{X_{1i}}} + \beta_2 \underset{\text{time}}{\underset{\uparrow}{X_{2i}}} + \epsilon_i$$

An "equivalent" model is

$$y_i = \alpha_0 + \beta_1(X_{1i} - \bar{X}_{1.}) + \beta_2(X_{2i} - \bar{X}_{2.}) + \epsilon_i$$

For the first model:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{21} \\ 1 & X_{12} & X_{22} \\ 1 & X_{13} & X_{23} \\ 1 & X_{14} & X_{24} \\ 1 & X_{15} & X_{25} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

For the second model:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} \begin{bmatrix} 1 & X_{11} - \bar{X}_1 & X_{21} - \bar{X}_2 \\ 1 & X_{12} - \bar{X}_1 & X_{22} - \bar{X}_2 \\ 1 & X_{13} - \bar{X}_1 & X_{23} - \bar{X}_2 \\ 1 & X_{14} - \bar{X}_1 & X_{24} - \bar{X}_2 \\ 1 & X_{15} - \bar{X}_1 & X_{25} - \bar{X}_2 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix} = \mathbf{W}\boldsymbol{\gamma} + \boldsymbol{\epsilon}$$

The space spanned by the columns of X is the same as the space spanned by the columns of $\mathbf{W}$. Find matrices $\mathbf{G}$ and F such that $X = \mathbf{WG}$ and $\mathbf{W} = XF$.

Definition 9.
Consider two linear models:

1. $E(\mathbf{y}) = X\boldsymbol{\beta}$ and $V(\mathbf{y}) = \boldsymbol{\Sigma}$ and,

2. $E(\mathbf{y}) = \mathbf{W}\boldsymbol{\gamma}$ and $V(\mathbf{y}) = \boldsymbol{\Sigma}$

where X is an $n \times k$ model matrix and $\mathbf{W}$ is an $n \times q$ model matrix.
We say that one model is a **reparameterization** of the other if there is a $k \times q$ matrix F and a $q \times k$ matrix $\mathbf{G}$ such that

$$\mathbf{W} = XF \text{ and } X = \mathbf{WG}.$$

The previous examples illustrate that if one model is a reparameterization of the other, then

(i) $\text{rank}(X) = \text{rank}(\mathbf{W})$

(ii) Least squares estimates of the response means are the same, i.e., $\hat{\mathbf{y}} = P_X\mathbf{y} = P_{\mathbf{W}}\mathbf{y}$

(iii) Residuals are the same, i.e.,

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = (I - P_X)\mathbf{y} = (I - P_w)\mathbf{y}$$

(iv) An unbiased estimator for σ^2 is provided by

$$MSE = SSE/(n - rank(X))$$

where,

$$\begin{aligned} \text{SSE} = \mathbf{e}^T\mathbf{e} &= \mathbf{y}^T(I - P_X)\mathbf{y} \\ &= \mathbf{y}^T(I - P_{\mathbf{W}})\mathbf{y} \end{aligned}$$

Reasons for reparameterizing models:

(i) Reduce the number of parameters
- Obtain a full rank model
- Avoid use of generalized inverses

(ii) Make computations easier
- In the previous examples, $\mathbf{W}^T\mathbf{W}$ is a diagonal matrix and $(\mathbf{W}^T\mathbf{W})^{-1}$ is easy to compute

(iii) More meaningfull interpretation of parameters.

Result 10.

Suppose two linear models,

$$(1)\quad E(\mathbf{y}) = X\boldsymbol{\beta} \quad V(\mathbf{y}) = \boldsymbol{\Sigma}$$

and

$$(2)\quad E(\mathbf{y}) = \mathbf{W}_{-}\gamma \quad V(\mathbf{y}) = \boldsymbol{\Sigma}$$

are reparameterizations of each other, and let F be a matrix such that $\mathbf{W} = XF$. Then

(i) If $\mathbf{C}^T\boldsymbol{\beta}$ is estimable for the first model, then $\boldsymbol{\beta} = F\boldsymbol{\gamma}$ and $\mathbf{C}^T F\boldsymbol{\gamma}$ is estimable under Model 2.

(ii) Let $\hat{\boldsymbol{\beta}} = (X^TX)^{-}X^T\mathbf{y}$ and $\hat{\boldsymbol{\gamma}} = (\mathbf{W}^T\mathbf{W})^{-}\mathbf{W}^T\mathbf{y}$. If $\mathbf{C}^T\boldsymbol{\beta}$ is estimable, then

$$\mathbf{C}^T\hat{\boldsymbol{\beta}} = \mathbf{C}^T F\hat{\boldsymbol{\gamma}}$$

.

.

Example 23.

Consider an experiment with 10 batteries. Two types of plate material (labeled 1 and 2)were randomly assigned to the 10 batteris using a balanced and completely randomized design. Temperatures of $x_1 = 0°C$, $x_2 = 30°C$, $x_3 = 60°C$, $x_4 = 90°C$ and $x_5 = 120°C$ were randomly assigned to 2 batteries with plate material type 1 and 2 respectively. Consider a Gauss-Markov model

$$y_{ij} = \mu + \alpha_i + \gamma X_j + \epsilon_{ij} \qquad \text{Model (1)}$$

where

- y_{ij} is the battery life time for the battery assigned to the j^{th} level of the temperature and the i^{th} level of the battery type,
- X_j denote the level of temperature administered to the battery,
- $\mu, \alpha_1, \alpha_2, \gamma$ are unknown parameters, and
- ϵ_{ij} denotes a random error with $\epsilon_{ij} \sim NID(0, \sigma^2)$ where $\sigma^2 > 0$.

Use this model to answer the following questions. (You may express your answers in matrix notation, but define any new notation that you introduce.)

(a) Let $\boldsymbol{\beta} = (\mu, \alpha_1, \alpha_2, \gamma)^T$, $\mathbf{y} = [y_{11}, y_{12}, y_{13}, y_{14}, y_{15}, y_{21}, y_{22}, y_{23},$ and

$\boldsymbol{\epsilon} = [\epsilon_{11}, \epsilon_{12}, \epsilon_{13}, \epsilon_{14}, \epsilon_{15}, \epsilon_{21}, \epsilon_{22}, \epsilon_{23}, \epsilon_{24}, \epsilon_{25}]^T$. Model (1) can be expressed in the form $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. Do not impose any restriction on the parameters, write down matrix of $\mathbf{X}$ in kronecker form.

(b) Show that $\mu + \alpha_2 + \gamma x$ is estimable for any $x \in \mathbf{R}$.

(c) Show that any two matrices $\mathbf{W}$ and $\mathbf{X}$ have the same column space if there exist matrices $\mathbf{F}$ and $\mathbf{G}$ such that $\mathbf{WG} = \mathbf{X}$ and $\mathbf{XF} = \mathbf{W}$.

(d) Show that $\mathbf{X}$ in part (a) has the same column space as

$$\mathbf{W} = \begin{bmatrix} 1 & -1 & -2 \\ 1 & -1 & -1 \\ 1 & -1 & 0 \\ 1 & -1 & 1 \\ 1 & -1 & 2 \\ 1 & 1 & -2 \\ 1 & 1 & -1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix}.$$

2.11 Restrictions (side conditions)

- Give meaning to individual parameters
- Make individual parameters estimable
- Create a full rank model matrix
- Avoid the use of generalized inverses
- Restrictions must involve "non-estimable" quantities for the unrestricted "effects" model.

Example 24. An effects model

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

This model can be expressed as

$$\begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1,n_1} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2,n_2} \\ y_{31} \\ y_{32} \\ \vdots \\ y_{3,n_3} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{1,n_1} \\ \epsilon_{21} \\ \epsilon_{22} \\ \vdots \\ \epsilon_{2,n_2} \\ \epsilon_{31} \\ \epsilon_{32} \\ \vdots \\ \epsilon_{3,n_3} \end{bmatrix}$$

Impose the restriction

$$\alpha_3 = 0$$

Then, $E(y_{1j}) =$

$E(y_{2j}) =$
$E(y_{3j}) =$
and

$$\begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1,n_1} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2,n_2} \\ y_{31} \\ y_{32} \\ \vdots \\ y_{3,n_3} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{1,n_1} \\ \epsilon_{21} \\ \epsilon_{22} \\ \vdots \\ \epsilon_{2,n_2} \\ \epsilon_{31} \\ \epsilon_{32} \\ \vdots \\ \epsilon_{3,n_3} \end{bmatrix}$$

Write this model as $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$ where

$$X = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \end{bmatrix} \text{ and } \boldsymbol{\beta} = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{bmatrix}$$

Then, $X^T X =$
and

$X^T\mathbf{y} =$

and the unique OLS estimator for $\boldsymbol{\beta} = (\mu\ \alpha_1\ \alpha_2)^T$ is

Example 25.
Consider the model $y_{ij} = \mu+\alpha_i+\epsilon_{ij}$ with the restriction $\alpha_1+\alpha_2+\alpha_3 = 0$. Then, $\alpha_3 = -\alpha_1-\alpha_2$ and

$E(y_{1j}) =$
$E(y_{2j}) =$
$E(y_{3j} =$ and

$$\begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1,n_1} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2,n_2} \\ y_{31} \\ y_{32} \\ \vdots \\ y_{3,n_3} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \\ \vdots & \vdots & \vdots \\ 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{1,n_1} \\ \epsilon_{21} \\ \epsilon_{22} \\ \vdots \\ \epsilon_{2,n_2} \\ \epsilon_{31} \\ \epsilon_{32} \\ \vdots \\ \epsilon_{3,n_3} \end{bmatrix}$$

This model is $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$

The unique OLS estimator for $\boldsymbol{\beta} = (\mu\ \alpha_1\ \alpha_2)^T$ is

Example 26.
Consider the model $y_{ij} = \mu + \alpha_i + \epsilon_{ij}$ with the restriction $\alpha_1 = 0$. Then,

$E(y_{1j}) =$
$E(y_{2j}) =$
$E(y_{3j})$
and

$$\begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1,n_1} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2,n_2} \\ y_{31} \\ y_{32} \\ \vdots \\ y_{3,n_3} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{1,n_1} \\ \epsilon_{21} \\ \epsilon_{22} \\ \vdots \\ \epsilon_{2,n_2} \\ \epsilon_{31} \\ \epsilon_{32} \\ \vdots \\ \epsilon_{3,n_3} \end{bmatrix}$$

This model is $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$, with

$$X = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \end{bmatrix} \text{ and } \boldsymbol{\beta} = \begin{bmatrix} \mu \\ \alpha_2 \\ \alpha_3 \end{bmatrix}$$

The unique OLS estimator for $\boldsymbol{\beta} = (\mu\ \alpha_1\ \alpha_2)^T$ is

MEME16203 LINEAR MODELS©Dr Yong Chin Khian

The restrictions (i.e. the choice of one particular solution to the normal equations) have no effect on the OLS estimates of estimable quantities. The estimated treatment means are:

$$E(\hat{y_{1j}}) = \hat{\mu} = \bar{y}_{1.}$$

$$E(\hat{y_{2j}}) = \hat{\mu} + \hat{\alpha}_2 = \bar{y}_{2.}$$

$$E(\hat{y_{3j}}) = \hat{\mu} + \hat{\alpha}_3 = \bar{y}_{3.}$$